```python
# Import necessary libraries
!pip install seaborn
!pip install scikit-learn

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error
```

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: seaborn in /home/immr/.local/lib/python3.13/si
te-packages (0.13.2)
Requirement already satisfied: numpy!=1.24.0,>=1.20 in /usr/lib64/python3.13/
site-packages (from seaborn) (2.2.6)
Requirement already satisfied: pandas>=1.2 in /home/immr/.local/lib/python3.1
3/site-packages (from seaborn) (2.2.3)
Requirement already satisfied: matplotlib!=3.6.1,>=3.4 in /home/immr/.local/l
ib/python3.13/site-packages (from seaborn) (3.10.3)
Requirement already satisfied: contourpy>=1.0.1 in /home/immr/.local/lib/pyth
on3.13/site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (1.3.2)
Requirement already satisfied: cycler>=0.10 in /home/immr/.local/lib/python3.
13/site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in /home/immr/.local/lib/pyt
hon3.13/site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (4.56.0)
Requirement already satisfied: kiwisolver>=1.3.1 in /home/immr/.local/lib/pyt
hon3.13/site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (1.4.8)
Requirement already satisfied: packaging>=20.0 in /usr/lib/python3.13/site-pa
ckages (from matplotlib!=3.6.1,>=3.4->seaborn) (24.2)
Requirement already satisfied: pillow>=8 in /usr/lib64/python3.13/site-packag
es (from matplotlib!=3.6.1,>=3.4->seaborn) (11.1.0)
Requirement already satisfied: pyparsing>=2.3.1 in /home/immr/.local/lib/pyth
on3.13/site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (3.2.3)
Requirement already satisfied: python-dateutil>=2.7 in /usr/lib/python3.13/si
te-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /home/immr/.local/lib/python3.
13/site-packages (from pandas>=1.2->seaborn) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in /home/immr/.local/lib/python
3.13/site-packages (from pandas>=1.2->seaborn) (2025.2)
Requirement already satisfied: six>=1.5 in /usr/lib/python3.13/site-packages
(from python-dateutil>=2.7->matplotlib!=3.6.1,>=3.4->seaborn) (1.17.0)
Defaulting to user installation because normal site-packages is not writeable
Collecting scikit-learn
  Downloading scikit_learn-1.7.0-cp313-cp313-manylinux_2_17_x86_64.manylinux2
014_x86_64.whl.metadata (17 kB)
Requirement already satisfied: numpy>=1.22.0 in /usr/lib64/python3.13/site-pa
ckages (from scikit-learn) (2.2.6)
Collecting scipy>=1.8.0 (from scikit-learn)
  Downloading scipy-1.15.3-cp313-cp313-manylinux_2_17_x86_64.manylinux2014_x8
6_64.whl.metadata (61 kB)
Collecting joblib>=1.2.0 (from scikit-learn)
  Downloading joblib-1.5.1-py3-none-any.whl.metadata (5.6 kB)
Collecting threadpoolctl>=3.1.0 (from scikit-learn)
  Downloading threadpoolctl-3.6.0-py3-none-any.whl.metadata (13 kB)
Downloading scikit_learn-1.7.0-cp313-cp313-manylinux_2_17_x86_64.manylinux201
4_x86_64.whl (12.5 MB)
                                        ━━━━━━━━━━━━━━━━━━━━━━━ 12.5/12.5 MB 5.6 MB/s eta 0:00:00
m eta 0:00:010:01:01
Downloading joblib-1.5.1-py3-none-any.whl (307 kB)
Downloading scipy-1.15.3-cp313-cp313-manylinux_2_17_x86_64.manylinux2014_x86_
64.whl (37.3 MB)
                                        ━━━━━━━━━━━━━━━━━━━━━━━ 37.3/37.3 MB 5.8 MB/s eta 0:00:00
m eta 0:00:01[36m0:00:01
Downloading threadpoolctl-3.6.0-py3-none-any.whl (18 kB)
Installing collected packages: threadpoolctl, scipy, joblib, scikit-learn
```

```
Successfully installed joblib-1.5.1 scikit-learn-1.7.0 scipy-1.15.3 threadpoo
lctl-3.6.0
```

In [13]:
```python
warnings.filterwarnings('ignore')
```

In [14]:
```python
# Generate synthetic dataset for building features and energy efficiency rat
np.random.seed(0)
data_size = 500
data = {
    'WallArea': np.random.randint(200, 400, data_size),
    'RoofArea': np.random.randint(100, 200, data_size),
    'OverallHeight': np.random.uniform(3, 10, data_size),
    'GlazingArea': np.random.uniform(0, 1, data_size),
    'EnergyEfficiency': np.random.uniform(10, 50, data_size)  # Energy effic
}
df = pd.DataFrame(data)
```

In [15]:
```python
# Data preprocessing
X = df.drop('EnergyEfficiency', axis=1)
y = df['EnergyEfficiency']
```

In [16]:
```python
# Visualize the relationships between features and the target variable (Ener
sns.pairplot(df, x_vars=['WallArea', 'RoofArea', 'OverallHeight', 'GlazingAr
plt.show()
```
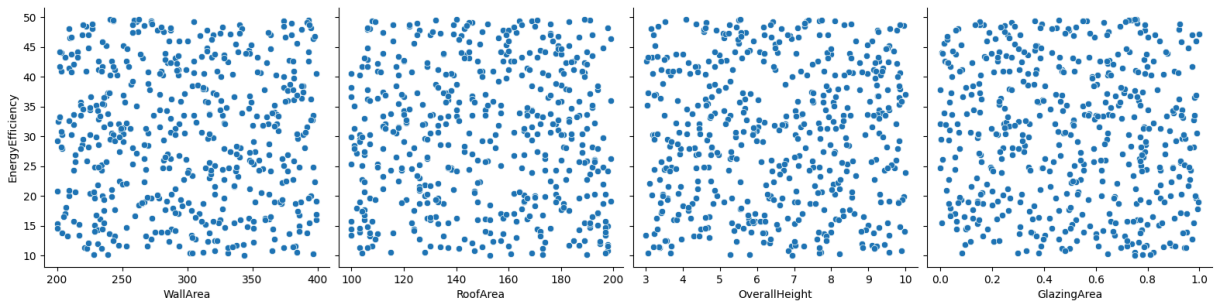

In [17]:
```python
# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, ran
```

In [18]:
```python
# Train a Random Forest model
model = RandomForestRegressor()
model.fit(X_train, y_train)
```

Out[18]:
```
▼ RandomForestRegressor    ⓘ ⍰

▶ Parameters
```

In [19]:
```python
# Predict and evaluate
predictions = model.predict(X_test)
mse = mean_squared_error(y_test, predictions)
print(f"Mean Squared Error: {mse}")
```

```
Mean Squared Error: 146.02051889440582
```

```
In [20]:  # Plot the True values vs Predicted values
          plt.figure(figsize=(10, 6))
          plt.scatter(y_test, predictions)
          plt.xlabel("True Values")
          plt.ylabel("Predictions")
          plt.title("True Values vs Predicted Values")
          plt.plot([y_test.min(), y_test.max()], [y_test.min(), y_test.max()], 'k--')
          plt.show()
```